Automatic Laser Dot Detection and Tracking in Underwater Video Streams Using Color-based Segmentation and Wavelet-based Template Matching

Fitzgerald J. Archibald
University of California, Santa Barbara

***Abstract*—Underwater videos are made with two red laser dots of fixed separation distance and visible in captured frame. Red laser dots in captured frame are used for distance resolution estimation in order to associate statistics with area. This paper proposes a method for automatic detection of the dots using wavelet sub-band template matching and laser dot tracking. Segmentation based on color information and 8-connectivity are used as preprocessing steps to improve detection accuracy and optimize computational cycles of template matching stage. Template matching in the Stationary Wavelet Transform (SWT) domain is used for further isolating red laser dots. Horizontal sub-band at 3rd level decomposition using Haar basis function is used to eliminate unwanted details in template matching. Finally, a distance error minimization with constrained feature is used in detecting the red laser dots. This tracking step minimizes the training requirement of template matching stage whilst improving detection rate in the undersea video.**

***Index Terms*—Object detection, object tracking, template matching**

I. Introduction

Undersea videos are taken with two laser light sources mounted along the sides of the video camera. The laser dots reflected on sea floor are captured along with underwater objects of interest. The distance between the dots is fixed at a pre-defined value (say 10 cm). This distance measure is used in estimating area of the captured video frame. The area is required in certain biological or marine science studies to generate statistics of marine life.

In certain statistics gathering scenarios, the video frames are extracted into spatially non-overlapping video sequence. Spatial non-overlap is required to avoid counting same object twice while collecting statistics. Assuming the submarine (mounted with the video camera) moves at a constant pace, the video frames can be periodically dropped to produce spatially non-overlapping video sequence. Other methods, like global motion estimation can be used to estimate frame displacement and to create non-overlapping video sequence.

Manuscript received March 12, 2009.

F. J. Archibald is a Graduate student in the Electrical and Computer Engineering Department of University of California, Santa Barbara, CA 93117 USA (e-mail: fja@umail.ucsb.edu).

Once spatially non-overlapping video sequence is available, the next step is to estimate the area of the video frame. In order to estimate area, the distance between the laser dots need to be measured. It can be assumed that the inclination of the camera (say 35 degrees) is known for this estimation process. In this paper, a method for automatically detecting the red laser dots is proposed. Using the coordinates of the detected red laser dots, the distance between dots is estimated for computing area.

There are several challenges associated with laser dot detection in the spatially non-overlapping video sequence. The red dots vary in size, shape, and color. This is due to change in camera angle, depth and variation of submarine, changes in the property of the medium (water containing particles and micro-organism). The patterns in sea floor may have the spectral and spatial characteristics of the red dots. Sea floor contains several materials like mica and sand which reflect light similar to red dots. In some cases, the laser dots are occluded or scattered. Some marine animals also contain patterns similar to the red dots. For example, eyes of shrimp are red in color. Shells of planktons create white dots with similar intensity distribution characteristics as that of red laser dots. In addition, the video frame size is 1440x1080 pixels. This implies large amount of data for processing and thus requires faster algorithms.

Fig. 1 and Fig. 2 provide insight into the above mentioned challenges. In Fig. 1, the patterns of sea floor have characteristics similar to red laser dots. The fishes in Fig. 2 have patterns similar to red laser dots.

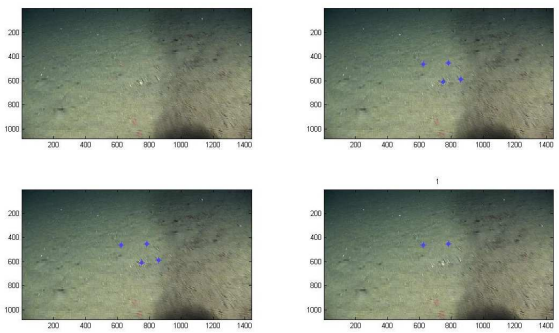

Fig. 1. Laser dot detection (Example 1)

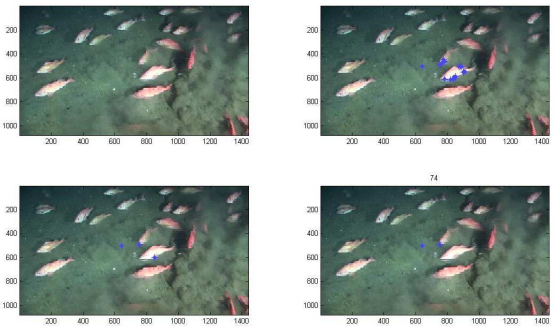

Fig. 2. Laser dot detection (Example 2)

On detection of dots, the coordinates are stored in an XML file which is taken as input by an annotation software or tool. At this point, a biologist can start annotating and collecting statistics using the spatially non-overlapping video frames.

Quellec *et al*. proposed an automatic method to detect micro-aneurysms in retinal photographs [1]. In their paper, multiple sub-bands of wavelet transformed images are used in template matching. The paper discusses choice of optimal wavelet transform using lifting scheme for the detection. A related approach of using Gaussian function as template was proposed to detect pulmonary modules in Helical Computer Tomography images [6].

Spencer *et al*. proposed a bilinear top-hat transformation and matched filtering to provide an initial segmentation of the images. This processed image is thresholded to form a binary image containing candidate microaneurysms. A region-growing algorithm is used to delineate each marked object and subsequent analysis of the size, shape, and energy characteristics of each candidate results in the final segmentation of microaneurysms [7].

Grisan and Ruggeri proposed local thresholding followed by an evaluation of a measure of the spatial density of the pixels selected at the first step for detection of hemorrhagic (dark) lesions in retinal images [4]. Image contrast normalization is used to improve the ability to distinguish between microaneurysms and other dots that occur on the retina [3].

In this paper, a color-based segmentation followed by 3x3 median filtering and 8-connected component segmentation is used on a region of interest (ROI), which will be in the vicinity of red laser dots, to detect potential dots. Further, the dimensions of the segmented objects are used for localization of the dots. The potential dot centers along with new ROI are provided as input for template matching. Stationary Wavelet Transform (SWT) is used to decompose the red channel of the ROI. The horizontal details on 3rd level decomposition is used for template matching. Sum of Squared Errors (SSE) is used as the cost function for template matching. The dots are accepted or rejected in template matching stage using thresholding. The thresholds are obtained using 700 dots training set. Further, the distance error of the dots in current frame with respect to previous frame is minimized while constraining minimum SSE. If the sum of SSE minima of the detected dots and distance error achieve minimum, the dots are classified as red laser dots. Otherwise, the dots with minimum distance error are classified as red laser dots.

The paper is organized into two main sections: design and results. Design is organized into segmentation, template matching, and tracking sub-sections. In results section, performance data and pros/cons of the method are brought out.

II. Design

In this section, the top level design is followed by a section on segmentation using color information and 8-connected component labeling. The next section, template matching, consists of wavelet decomposition, template matching, feature extraction, and feature matching sub-sections. Finally, a section on tracking-based dot selection/rejection is presented.

Since the red dot is positioned slightly above the center of the image, based on training data, a ROI is established around the red laser dots. The region is segmented using color information and 8-connected component labeling. The dot positions along with the R-channel data in the ROI is passed to template matching stage. The template matching stage decomposes the signal using SWT with “Haar” basis functions. Only the horizontal detail coefficients on 3rd level decomposition are retained for further processing.

The template sub-band is matched around the potential dot positions, provided by segmentation module. The features of the curves resulting from the template matching cost function (SSE) are extracted. The features include minimum, variance, energy around the minimum point and histogram of the cost function (SSE) output curve. These features are passed onto feature matching stage along with the dot positions from the segmentation stage. The feature matching stage uses thresholds from training data to reject or retain dot candidates from segmentation stage. The processing flow of above description is shown in Fig. 3.

A. Segmentation

Since the camera is at an inclination with respect to the sea floor, the red dots are shifted to the top half of the video frame.

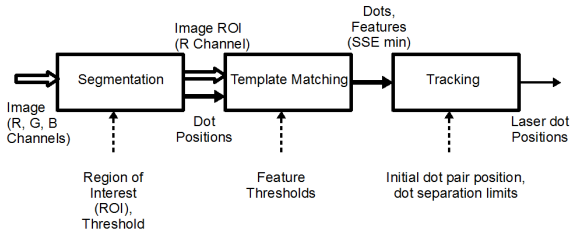

Fig. 3. Top level block diagram

In addition, the dots are most likely on either side of the vertical half line. This allows to create a smaller ROI, for further processing, from the larger input video frame. The segmentation process flowchart is shown in Fig. 4.

The maximum red component of all the pixels within the ROI is scaled, with a pre-defined scalar obtained from training, to derive red color threshold value. If the red color component of the pixels is greater than the threshold and red component of the pixel is greater than the blue and green components, the pixel is assigned a binary value of '1'. Otherwise, a value of '0' is assigned. Thus a binary map (image) of the ROI is created. This binary image is passed

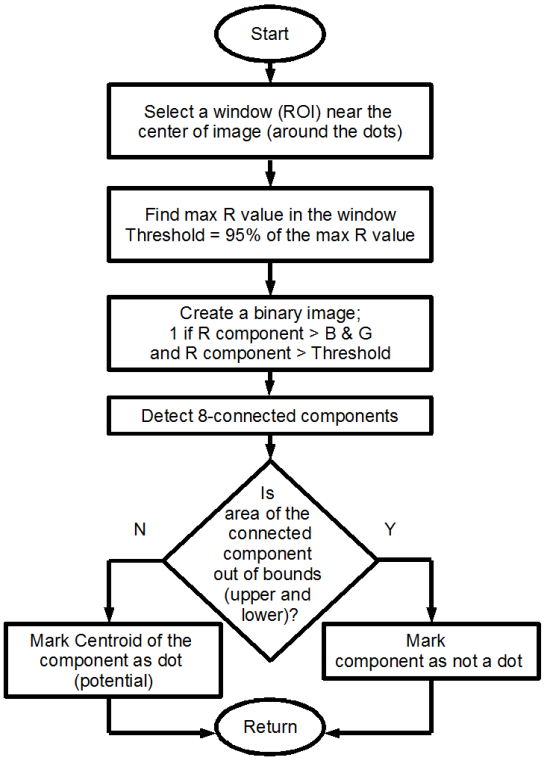

Fig. 4. Segmentation flowchart

through a 3x3 median filter to remove any speckles from segmentation.

The median filtered binary image is further segmented using 8-connected component labeling [5]. The 8-connected component labeling performs well for spherical isolated dots. Hence, 8-connected component labeling is preferred over 4-connected component labeling. Segmentation of the samples shown in Fig. 1 and Fig. 2 are provided in Fig. 5 and Fig. 6 respectively. Notice the dots in second example are not visible without careful examination. If the area of the connected component is larger than a threshold (maximum size of red laser dot obtained from training), the component is excluded from being a potential dot. The remaining components are detected as dots and their centroids are computed. A new ROI is established around the detected dots to reduce wavelet transformation dimension in next stage processing. The new ROI has tighter bounds as compared to the ROI used in segmentation. The dot positions along with the red color plane

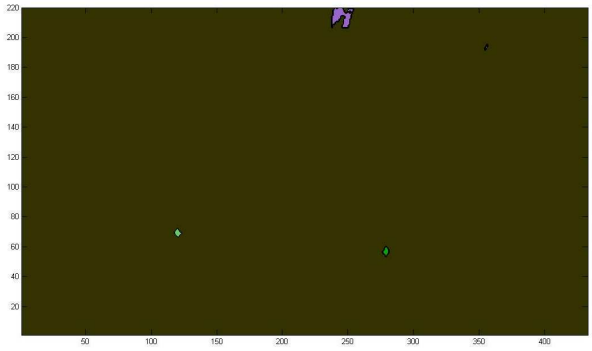

Fig. 5. Segmentation (Example 1)

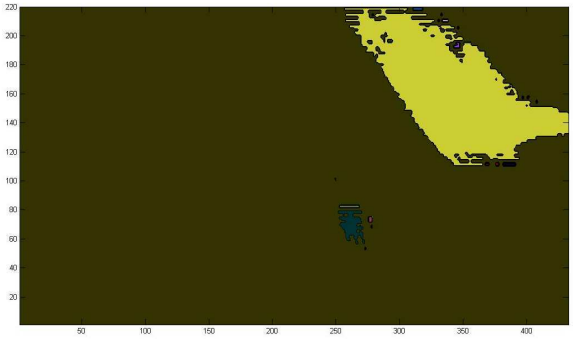

Fig. 6. Segmentation (Example 2)

of the ROI are passed as input to template matching stage for detecting red laser dots.

B. Template Matching

In template matching, a template is matched with the input image at the dot positions from segmentation. The template matching is implemented using sub-bands from wavelet decomposition. The wavelet decomposition is restricted to red channel data since the goal is to detect red laser dots whilst maximizing execution speed. However, all three channels (R, G, B) may be used to further refine the feature vectors. The block diagram of the template matching process is shown in Fig. 7.

The result of template matching is output 2-D curve of cost function, which is SSE in this case. The curve features are

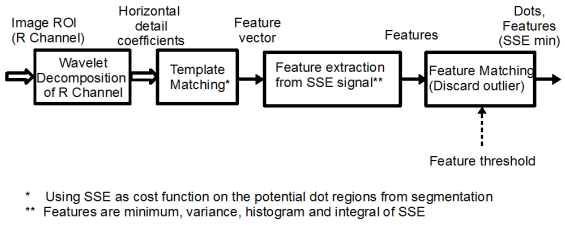

Fig. 7. Block diagram of template matching

extracted and matched with thresholds obtained from training data to retain or reject detected dots from segmentation.

This section is organized into four sub-sections: wavelet decomposition, template matching submodule, feature extraction and feature matching.

1) Wavelet Decomposition

The sub-band resolution needs to be retained since the laser dots are small in dimensions, the maximum being 20x24 pixels. Stationary wavelets are shift-invariant and maintain resolution, same as image size, of sub-bands [2]. Hence, a SWT decomposition of 3 levels is used in template matching module as shown in Fig. 8. Fig. 9 and Fig. 10 illustrate advantages of SWT over DWT with an example video frame containing red laser dots. The SWT retains the red laser dot in horizontal detail sub-band on 3rd level decomposition whereas the red laser dot is faintly visible in DWT case. More details on DWT can be found in [5].

The training data set shows a clear red dot isolation at 3rd level of decomposition using SWT. In the first two stages of decomposition, high frequency details are eliminated. The horizontal detail coefficients retain the red laser dots since the video frame is interlaced and not progressive. That is, the red

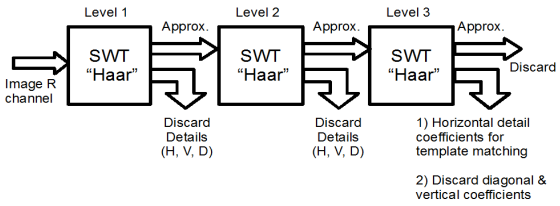

Fig. 8. Wavelet decomposition block diagram

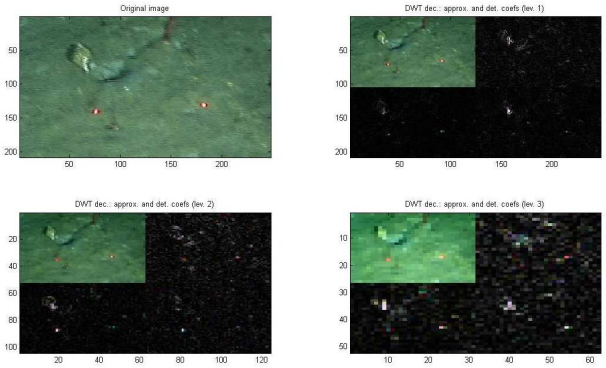

Fig. 9. DWT example

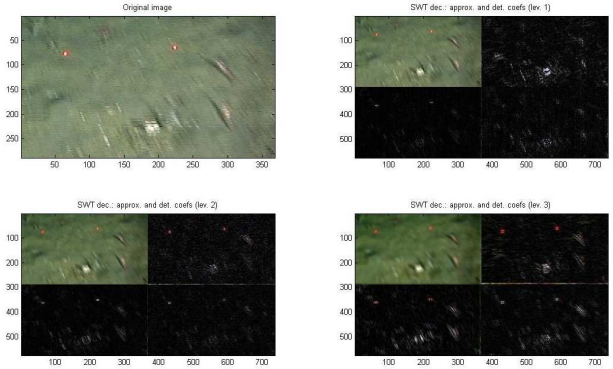

Fig. 10. SWT example

dots have higher vertical frequency than horizontal frequency. The same reasoning holds good for diagonal sub-band, which contains noise and high frequency details. By discarding the approximation sub-band, the variations in illumination at the recording time are eliminated.

Haar basis is well suited for sub-band analysis of circular edges as in red dots. The results are confirmed with training data using different wavelet filters namely Daubechies, Bi-orthogonal, Symlet, Coiflet, Meyer and Reverse Bi-orthogonal. Therefore, Haar basis functions are used for wavelet transformation.

2) Template matching Submodule

SSE is preferred over correlation and normalized correlation since the resulting curve is smoother and aids in curve fitting. Thus, SSE is used as cost function in template matching. A single template of a dot whose 3rd level horizontal sub-band, generated from SWT using Haar basis, is used for template matching. The template is of dimension 15x17 coefficients. The template is slid around the dot centers, available from segmentation, in both x and y directions to compute SSE. The search area is 42x44 coefficients. Since the dots are small enough, a single template is sufficient. Moreover, there is no need to detect the center of the dot, as the dot centers are available from segmentation stage. Fig. 11 shows the SSE output curves of the image in Fig. 1. The SSE curves 1 and 3 correspond to a red dot whereas 2 and 4 are that of white dots

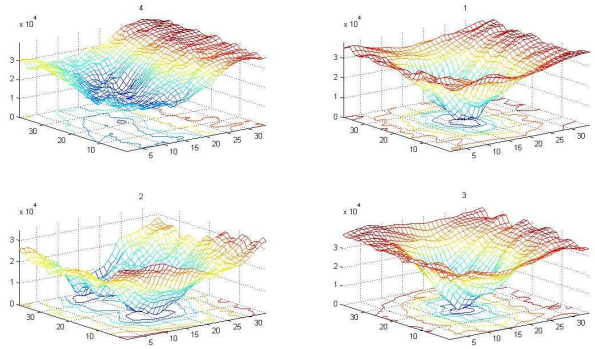

Fig. 11. Template matching (SSE curve examples)

from sea floor. The SSE curve features are extracted in next stage called feature extraction submodule.

3) Feature extraction

In this stage, the features of the SSE curves in the search window for each potential dot are extracted. The features are minimum SSE, variance along x-direction, variance along y-direction, 24x24 window integral of SSE curve centered around the minimum point of SSE, and histogram of SSE.

In case of potential dots, histogram will have a monotonically increasing bin height with increasing value of SSE as shown in Fig. 12. This is intuitive from the shape of the SSE curve (inverted bivariate Gaussian). The bin corresponding to the highest SSE value may not increase

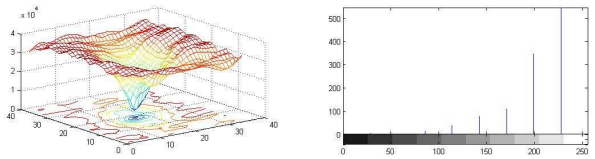

Fig. 12. Histogram of SSE (matching case)

monotonically, since the curve distribution for highest value of SSE depends on the window size and the size of template. On the contrary, SSE of false dots detected by segmentation stage produce histograms, which are non-monotonically increasing, as shown in Fig. 13. So, the number of monotonically increasing bins is used as a feature. In addition, observations from training data show that 3 to 4 bins (in 10 bin histograms)

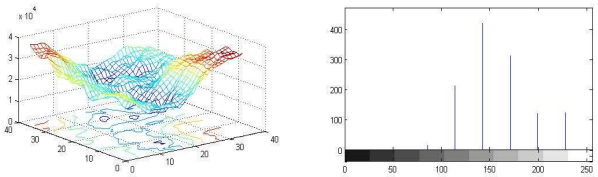

Fig. 13. Histogram of SSE (non-matching case)

contain more than 80% of the points in SSE curve. This can be visualized from the nature of the SSE curve. Hence, ratio of SSE amplitude distribution is used as an additional feature.

Once features are extracted, the features are passed to feature matching submodule.

4) Feature matching

When the red dots are distorted due to occlusion, scattering, or reflection from non-uniform surface (sea floor), the SSE curves may show similar shape as in plots 2 and 4 in Fig. 11. This implies that a simple inverted bivariate Gaussian approximation-based curve fitting of SSE may not yield good detection rate. However, the detection rate may be possibly increased with elaborate training.

Only a first level of dot detection or rejection is carried out using feature matching. That is, a wider threshold for the features obtained from training data are used for rejecting dots from segmentation. The thresholds are upper and lower limits for variance along x and y directions, deviation between the ratio of x and y variances and SSE integral. In addition, lower bounds of monotonically increasing bin count of SSE histogram, and ratio of amplitude distribution of SSE are used in this submodule. The remaining dots are passed to the laser dot tracking module.

C. Laser Dot Tracking

The number of features that can be used for detection of dots are smaller to conclude positive detection at template matching stage. In addition, patterns in sea floor and fishes may exhibit similar features as that of laser dots. So, the tracking stage is required for improving detection rate by minimizing dependency on training effectiveness.

The camera is maintained at a relatively constant angle between frames. Moreover, the depth of the camera with respect to the sea floor remains unchanged between frames. Hence, the laser dot position in previous frame can be used as a feature in identifying the red dots. Due to non-overlapping video frame extraction, the assumed constancy of angle and depth may be violated in some rare cases. However, these special cases may not affect detection rate if only two dots are detected at template matching stage.

Though bound checking of histogram was performed at template matching stage, a tighter threshold is established for the potential dot candidates in this stage. The tighter bound is required to avoid false laser dot detections.

The flowchart of the laser dot tracking algorithm is shown in Fig. 14. If exactly two dots are available from template matching stage, the dots correspond to red laser dots. Then, the previous red dot position is updated with the current red dot position, for use in next frame processing. If there is a dot or no dot detected from template matching stage, no red laser dots are detected as red laser dots have to occur in pair to be able to measure distance resolution.

If there are more than two dots from template matching stage, the vertical and horizontal distances between each dot are computed. The distance is populated in an upper triangular distance matrix as shown in step 1 of Fig. 14. This results in a matrix for x-separation and another for y-separation. The separation in y-direction is small (say 10 pixels), whereas separation in x-direction is large (say 250 pixels corresponding to 10 cm dot separation). Using learning data, upper threshold for separation in y-direction, and upper and lower thresholds for separation in x direction are established. The matrices containing x and y-separation are compared against the thresholds. The dot pairs that do not meet the separation distance criteria are unassociated. That is, the dots can not form a laser dot pair since the distance separation criteria is not met.

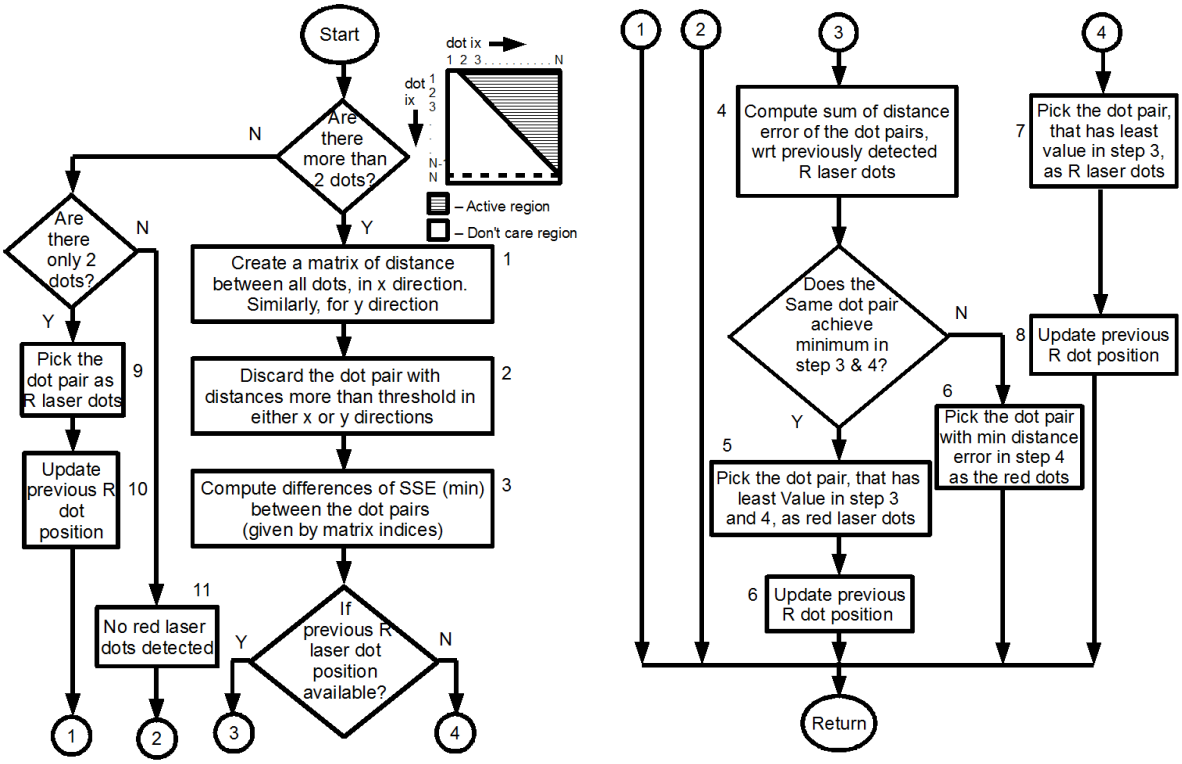

Fig. 14. Flowchart of tracking

The sum of the SSE minima for each valid dot pair is computed. The dots that match the template have a small SSE minimum at the matching position as compared to dot like patterns in the sea floor and sea animals. This is due to the fact that red dots are closer in size, shape, and illumination characteristics to the template than other natural patterns in the sea. Further, this observation is confirmed from training.

If previous frame laser dot position is not available, then the dot pair that has least value for sum of the SSE minima is detected as the red dot pair. Accordingly, the previous frame dot positions are updated with new red dot positions. This step is exercised during start up of the algorithm when there are no initial coordinates for red laser dots.

If previous laser dot position is available, the distance errors between current frame dot pairs and previous dot pair are computed. The dot pair that achieves minimum distance error and least SSE minimum sum is detected as laser dot pair. Also, previous red dot position is updated. If no red laser dots are detected, the dot pair achieving minimum distance error is selected as the red dot pair. Since this selection is not based on conclusive features, the previous red dot position is retained.

To improve the accuracy of this stage, a user fed dot position on start-up of algorithm is recommended. Fig. 15 shows the x and y coordinates of the red laser dots obtained without providing any initial conditions. The first row corresponds to x and y positions of the first laser dot and second row corresponds to the x and y positions of second laser dot. The y positions of the two dots exhibit opposing trends due to depth variations.

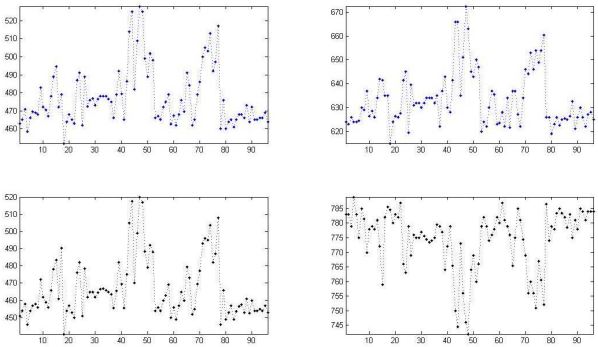

Fig. 15. Example of laser dot position tracking

III. Results

The algorithm is tested with 1080 video frames (11 sequences, each containing 99 spatially non-overlapping frames). The initial dot position is not provided and all the default setting from training data is used. The result is summarized in Table I. Negative result indicates either false positives or false negatives. If the dots are not visible to human eye or if the frame data is noisy to the extent it can not be used for statistics, the corresponding frames are counted in the columns “Bad frames”. Detection rate and false rate provide the detection success and failure percentages respectively. On average, the detection rate is 96% of the good frames. The remaining 4% is due to false positives and/or negatives in good frames.

TABLE I
Detection Performance

Stream*	Negative (frames)	Bad Frames (frames)	Total (frames)	Good Frames (frames)	Detection rate (%)	False rate (%)
B01C0704	5	21	99	73	94	6
B01C0705	5	29	99	65	93	7
B01C0706	0	10	99	89	100	0
B01C0708	6	14	99	79	93	7
B01C0709	4	21	99	74	95	5
B01C0710	2	13	99	84	98	2
B01C0801	0	17	99	82	100	0
B01C0802	7	9	99	83	92	8
B01C0803	1	3	99	95	99	1
B01C0804	4	33	99	62	94	6
B01C0805	0	53	99	46	100	0
Average	34	223	1089	832	96	4

* Files are sampled temporally at 3 sec rate to create spatially non-overlapping image sequence Dive 01 - 28 July 07 - DW6 - Guggenheim

Table II provides the execution profile of the algorithm. The majority of execution cycles (75%) is spent in SSE computation. The next highest cycle usage is by SWT (19%). The remaining processing steps use 6% of the total execution cycles. If needed, speed optimization can be achieved by using multi-step window sliding instead of single-step window sliding in SSE [1].

TABLE II
Execution Profile

Module	% Execution Time
SSE	74.70%
SWT	19.40%
Segmentation	3.40%
Feature Extraction	0.90%
Others	1.60%

The algorithm execution in Matlab consumed 0.58 sec of processing time per 1440x1080 resolution frame in a Intel Centrino core 2 Duo (T5750) at 2GHz with 3GB RAM.

IV. Conclusion

Template matching of subbands, using SSE cost functions, obtained from SWT decomposition provides sufficient features for matching. The SSE output is characterized by inverted bivariate asymmetric Gaussian function. Thereby, a loosely modeled asymmetric bivariate Gaussian feature matching coupled with distance tracking eliminates the necessity for extensive training. Moreover, the interlaced video simplified computational complexity by needing only 3rd level horizontal sub-bands for template matching.

A shift invariant wavelet transform is well suited for maintaining the resolution of the sub-bands on decomposition, which is desirable since the laser dots are small. In addition, Haar basis functions are sufficient for sub-band template matching for circular patterns like red laser dot. In order to speed up processing and improve detection rate in template matching, a preprocessing stage like color segmentation and connected components is used to identify potential dots and restrict ROI.

Due to the lack of sufficient features in red laser dots, similarity with features in natural patterns and high volume of video data, automatic detection method requires a tracking algorithm to enhance detection accuracy. The tracking algorithm is based on distance error minimization between

current frame dot pairs and previous frame dot pair with constrained least sum of SSE minima.

The limitation of this method is the possibility of initial tracking errors. However, this limitation can be overcome by setting initial dot coordinates during startup. In addition, the CPU load varies across frames depending on the number of potential dot candidates and ROI from segmentation stage. This is not a limiting factor in off-line computation for which the method is developed.

Acknowledgment

The author thanks Prof. B.S. Manjunath, Professor, Electrical and Computer Engineering (ECE) and Dr. Robert Miller, Post Doctoral Researcher, Marine Science Institute in University of California, Santa Barbara (UCSB) for providing technical inputs for the project. Also, the author thanks Dmitry Fedorov, Doctoral candidate, ECE, UCSB for providing technical insights, and software for frame extraction and annotation.

References

[1] G. Quellec, M. Lamard, P.M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal Wavelet Transform for the Detection of Microaneurysms in Retina Photographs," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1230-1241, Sept. 2008.

[2] G. P. Nason, and B.W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*, A. Antoniadis, and G. Oppenheim Eds. New York: Springer, Feb. 1995, pp. 281-300.

[3] A.D. Fleming, S. Philip, K.A. Goatman, J.A. Olson, and P.F. Sharp, "Automated microaneurysm detection using local constrast normalization and local vessel detection," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1223-1232, Sep. 2006.

[4] E. Grisan, and A. Ruggeri, "A hierarchical Bayesian classification for non-vascular lesions detection in fundus images," *Proceedings of the 3rd European Medical and Biological Engineering Conference*, Prague, Czech Republic, vol. 11, Nov. 2005.

[5] R.C. Gonzalez, and R.E. Woods, *Digital Image Processing* Delhi, India: Pearson Education, 2002, pp. 66-69, pp. 349-403.

[6] Y. Lee, and H. Hara, "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," *IEEE Transactions on Medical Imaging*, vol. 20, no. 7, pp. 595-604, Jul. 2001.

[7] T. Spencer, J. Olson, K. McHardy, P. Sharp, and J. Forrester, "An image-processing strategy for the segmentation and quantification in microaneurysms in fluorescein angiograms of the ocular fundus," *Computers and Biomedical Research*, vol. 29, no. 4, pp. 284-302, Aug. 1996.

Fitzgerald Archibald was born in Nagercoil, Tamilnadu, India in 1975. He earned his B.E in electronics and communication engineering from PSG College of Technology, Coimbatore, Tamilnadu, India in 1996. He is pursuing MS in electronics and computer engineering at University of California, Santa Barbara, USA.

He worked on control systems software development for geo-synchronous satellites from 1996 to 1999 in ISRO Satellite Centre, Bangalore, India. In 2001-2002, he worked on speech decoder, real-time kernel, and audio algorithms for DVD audio team in Sony Electronics, San Jose, USA. While in Philips Semiconductors (Bangalore, India, and Sunnyvale, USA) in 1999-2001 and 2002-2004, he worked on audio algorithms, codecs and systems software for set-top-box, analog/digital TV, and internet audio player. He is part of the DSPS group in Texas Instruments (TI) Inc, Bangalore, India from 2004-till date working on audio/speech/video/imaging algorithms, codecs and systems software development for portable audio/media players, voice recorders, digital still cameras, security cameras, and broad market software development kits. He has 9 pending US patent applications and published 13 technical papers. His interests are multimedia and control algorithms and systems.

Mr. Archibald received Manufacturing Incentive Award from TI for his work on Multi-Format Audio Player in Jul, 2008. He received Innovators in Action Award (2Q2008) from Texas Instruments for work on DM355 based IP Network Camera Reference Design. He is a member of AES Signal Processing committee.